

REMARKS

Reconsideration and withdrawal of the rejections set forth in the Final Office Action dated September 15, 2008, is respectfully requested in view of this amendment.

Claim 15 has been amended. Claim 13 has been cancelled without prejudice or disclaimer. In this regard, Applicants note that the amended claim merely clarifies the subject matter recited in the rejected claims, but does not narrow the scope of the claims. Claims 1-12 and 14-18 are pending in this application.

In the outstanding Office Action, the Examiner objected to claims 1 and 13 as being Duplicated Claims and rejected claims 1-18 under 35 U.S.C. §103(a) as being unpatentable over U.S. Patent Publication No. 2003/0065635 to Sahami et al. (hereinafter referred to as “the Sahami et al. ‘635 publication”) in view of U.S. Patent No. 5,787,422 to Tukey et al. (hereinafter referred to as “the Tukey et al. ‘422 patent”).

By this Response and Amendment, claim 15 has been amended to clarify the claimed subject matter intended by the Applicants. Support for the amendment can be found from page 18, lines 21-26 of the original specification. Responsive to the Examiner’s objection, claim 13 has been cancelled.

It is respectfully submitted that the above amendments introduce no new matter within the meaning of 35 U.S.C. §132.

Claim Objection

The Examiner objected to claims 1 and 13 as being Duplicated Claims.

Response

By this Response and Amendment, Claim 13 has been cancelled without prejudice or disclaimer to the contents therein. Applicants respectfully submit that the above amendment obviates the Examiner’s objection to the claims.

Claim Rejections under 35 U.S.C. §103(a)

The Examiner rejected claims 1-18 under 35 U.S.C. §103(a) as being unpatentable over the Sahami et al. '635 publication in view of the Tukey et al. '422 patent.

Response

By this Response and Amendment, Claim 13 has been cancelled without prejudice or disclaimer to the contents therein, thereby making the rejection thereto moot.

Applicants respectfully traverse the remaining rejections since all of the features of the presently claimed subject matter are not disclosed by the cited references.

In order to establish a *prima facie* case of obviousness, the Examiner must establish: (1) some suggestion or motivation to modify the references exists; (2) a reasonable expectation of success; and (3) the prior art references teach or suggest all of the claim limitations.

Applicants respectfully submit that the Tukey et al. '422 patent fails to cure the deficiencies of the Sahami et al. '635 publication with respect to the claimed subject matter in accordance with Applicants' independent claim 1 and further, does not suggest a teaching or motivation to reach such subject matter as claimed in the instant application. Applicants further respectfully submit that the cited prior art of record does not suggest a teaching or motivation to reach such subject matter as claimed in the instant application.

Applicant's independent claim 1 sets forth:

"A method of determining cluster attractors for a plurality of documents, each document comprising at least one term, each term comprising one or more words, the method comprising: calculating, in respect of each term, a probability distribution indicative of the frequency of occurrence of the, or each, other term that co-occurs with said term in at least one of said documents; calculating, in respect of each term, the entropy of the respective probability distribution; selecting at least one of said probability distributions as a cluster attractor depending on the respective entropy value."

Amended claim 15 has been discussed above and recites:

“A method of clustering a plurality of documents, each document comprising at least one term, each term comprising one or more words, the method comprising determining cluster attractors in accordance with the method of Claim 1; comparing each document with each cluster attractor; and assigning each document to one or more cluster attractors depending on the similarity between the document and the cluster attractors.”

Firstly, Applicants respectfully disagree with the Examiner’s assertion that the substance of its arguments concerning the Sahami et al. ‘635 publication is that the Sahami et al. ‘635 publication is concerned with structured data as opposed to unstructured data. Rather, the crux of the Applicants’ argument concerning the Sahami et al. ‘635 publication is that the Sahami et al. ‘635 publication does not use cluster attractors. This is explained in further detail hereinafter.

The Examiner has asserted that the Sahami et al. ‘635 publication discloses all of the features of claim 1 except the cluster attractors, and that each term comprises one or more words. Applicants agree that the Sahami et al. ‘635 publication does not disclose the cluster attractors or that each term comprises one or more words, but respectfully disagrees with regard to the other features of claim 1 for the reasons explained below.

As acknowledged by the Examiner, the Sahami et al. ‘635 publication does not disclose a method of determining cluster attractors. The Sahami et al. ‘635 publication does not use cluster attractors when clustering sets of data. Instead, the Sahami et al. ‘635 publication assigns attributes and attribute values to the data and then performs a statistical analysis of the attribute values to identify which attribute has the highest influence on the other attributes. A respective cluster is then established for each attribute value that the most influential attribute can take (see in particular the example provided in paragraph [0034] of the Sahami et al. ‘635 publication).

The Tukey et al. ‘422 patent does not disclose a method of determining cluster attractors, although it indicates (col. 10 lines 4-7) that “clusters are often defined by a set of attractors, each essentially a vector that summarizes the vectors of each document belonging to the cluster, e.g. a “centroid” of those vectors.”

It is respectfully submitted that one of ordinary skill in the art could not combine the teaching of Tukey et al. '422 patent regarding cluster attractors with those of the Sahami et al. '635 publication, since the Sahami et al. '635 publication does not use or need cluster attractors.

Claim 1 recites "A method of determining cluster attractors for a plurality of documents, *each document comprising at least one term, each term comprising one or more words...*". The Sahami et al. '635 publication does not disclose this feature. Instead, the clustering of the Sahami et al. '635 publication is performed on data sets such as a database, data warehouse or data mart (see Abstract and paragraph [0031] of the Sahami et al. '635 publication).

The Tukey et al. '422 patent does disclose documents having at least one term comprising one or more words. However, one of ordinary skill in the art would not combine the Tukey et al. '422 patent's teaching in this regard with the Sahami et al. '635 publication since the Sahami et al. '635 publication's teaching relates solely to assigning attributes and attribute values to data sets.

Claim 1 further recites "...*calculating, in respect of each term, a probability distribution indicative of the frequency of occurrence of the, or each, other term that co-occurs with said term in at least one of said documents...*". This feature of claim 1 is concerned with how many times different terms (e.g. words) appear together in documents. By way of example, in cases where a "term" is a single word, then the probability distribution may indicate the respective number of times other words appear together with a given word in at least one of the documents.

The Sahami et al. '635 publication does not disclose this feature because it is not concerned with documents comprising terms and so does not contemplate the co-occurrence of terms within documents. Instead, as previously discussed, the Sahami et al. '635 publication is concerned with attributes and attribute values assigned to data sets.

The Examiner has referred to paragraphs [0058] and [0059] of the Sahami et al. '635 publication concerning probabilistic algorithms. However, the Sahami et al. '635 publication does not disclose the calculation of probability distributions that are indicative of the frequency of occurrences of terms in the documents as stipulated by Claim 1. Instead, the Sahami et al.

‘635 publication uses conditional probability to evaluate the influence of an attribute given a cluster of data records (see paragraphs [0066] and [0068]).

The Examiner also refers to paragraph [0027] of the Sahami et al. ‘635 publication, which relates to the COBWEB clustering technique. COBWEB is presented as a prior art technique for performing clustering using tree structures and is not actually used by the Sahami et al. ‘635 publication. The COBWEB technique uses probability to assign data points to data clusters. For example, referring to Figure 4 of the Sahami et al. ‘635 publication, if a new record X equals (x1, x2...xN) has its first attribute value equal to x1, then with probability 1.0, this record should be assigned to cluster C₁. Hence, the probability techniques disclosed by COBWEB do not relate to the frequency of occurrence of terms (e.g. words) within documents.

The Tukey et al. ‘422 patent also does not disclose calculating, in respect of each term, a probability distribution indicative of the frequency of occurrence of the, or each, other term that co-occurs with said term in at least one of said documents. As discussed above, the Tukey et al. ‘422 patent does not disclose his own method of determining cluster attractors but does refer to a document clustering technique in which the cluster attractor is taken as the “centroid” of vectors representing the documents in a cluster (col. 10 lines 4-7 of the Tukey et al. ‘422 patent). Calculating the centroid of document vectors is a completely different approach to determining attractors than the calculation of probability distributions concerning the frequency of occurrence of terms within documents as defined by the features of claim 1.

Since neither the Sahami et al. ‘635 publication nor the Tukey et al. ‘422 patent disclose “*...calculating, in respect of each term, a probability distribution indicative of the frequency of occurrence of the, or each, other term that co-occurs with said term in at least one of said documents...*” as recited in claim 1, the combination of references fails to lead one of ordinary skill in the art to such features.

Claim 1 further recites “*...calculating, in respect of each term, the entropy of the respective probability distribution...*”.

The Sahami et al. ‘635 publication does not disclose calculating the entropy of the respective probability distributions. As indicated above, the Sahami et al. ‘635 publication does

not disclose the calculation of probability distributions as defined by Claim 1 and so it follows that it cannot disclose the calculation of the entropy of those probability distributions. The Sahami et al. '635 publication does disclose the use of entropy as part of its calculations in paragraphs [0083] to [0087], but only as a means of eliminating features that are not useful in identifying clusters – see in particular paragraphs [0084] to [0086].

The Tukey et al. '422 patent does not disclose the use of entropy in his calculations.

Since neither the Sahami et al. '635 publication nor the Tukey et al. '422 patent disclose "*...calculating, in respect of each term, the entropy of the respective probability distribution...*", their combined teachings fails to lead one of ordinary skill in the art to this feature of claim 1.

Claim 1 also recites "*...selecting at least one of said probability distributions as a cluster attractor depending on the respective entropy value.*"

The Sahami et al. '635 publication does not disclose this feature since, as indicated above, he does not use cluster attractors, nor does he calculate the probability distributions of claim 1, nor does he calculate the entropy of said probability distributions.

The Tukey et al. '422 patent does not disclose this feature since, as indicated above, he uses the centroid technique for cluster attractors, he does not calculate the probability distributions of claim 1, nor does he calculate the entropy of said probability distributions.

Since neither the Sahami et al. '635 publication nor the Tukey et al. '422 patent discloses selecting at least one of said probability distributions as a cluster attractor depending on the respective entropy value, their combined teachings could not lead one of ordinary skill in the art to this feature of claim 1.

In summary, while the Tukey et al. '422 patent does make reference to a document clustering method involving cluster attractors, this is meaningless in the context of the Sahami et al. '635 publication since the Sahami et al. '635 publication does not use cluster attractors.

Moreover, neither the Sahami et al. '635 publication nor the Tukey et al. '422 patent disclose the following features recited in claim 1: "*...calculating, in respect of each term, a probability distribution indicative of the frequency of occurrence of the, or each, other term that co-occurs with said term in at least one of said documents; calculating, in respect of each term,*

the entropy of the respective probability distribution; selecting at least one of said probability distributions as a cluster attractor depending on the respective entropy value.”

Therefore, Applicants respectfully submit that the combination of the Sahami et al. '635 publication and the Tukey et al. '422 patent fails to disclose, teach, or suggest the novel and unobvious features of “a method of determining cluster attractors for a plurality of documents, each document comprising at least one term, each term comprising one or more words, the method comprising: calculating, in respect of each term, a probability distribution indicative of the frequency of occurrence of the, or each, other term that co-occurs with said term in at least one of said documents; calculating, in respect of each term, the entropy of the respective probability distribution; selecting at least one of said probability distributions as a cluster attractor depending on the respective entropy value” as recited in claim 1 of the instant application.

Accordingly, Applicants submit that claim 1 is novel and unobvious over the prior art of record, and submits that previously presented claims 2-12, amended claim 15 and claims 16 and 17 which depend from amended dependent claim 15, are also then patentable over the prior art of record and request indication of such.

Notwithstanding the above, it is respectfully submitted that the claims depending from claim 1 also have features which are novel, unobvious and patentable *per se*.

With regard to the Examiner's comments on claim 2, paragraph [0012] of the Sahami et al. '635 publication discloses the K-means technique for cluster identification. The K-means technique does not determine cluster attractors using probability distributions relating to the co-occurrence of terms (e.g. words) in a document as defined in claim 2. Instead, the K-means technique simply selects a plurality of data points to serve as “centroids” and then adjusts the centroids in an iterative process using a distance measurement between the centroids and the surrounding data points (see in particular paragraph [0012] “The process starts with the placement of k centroids in the domain space. Then the centroids are adjusted in an iterative process until their position stabilizes..... The resulting clusters are formed by those data points within a certain distance of the centroids....”). Hence, the co-occurrence of terms within

document is not considered by the Sahami et al. '635 publication since it is not relevant to the K-means technique. It is also noted that the Sahami et al. '635 publication discloses K-means as prior art and does not actually use it itself – as explained above, the Sahami et al. '635 publication does not use centroids or cluster attractors. Accordingly, the features of claim 2 are not disclosed by the Sahami et al. '635 publication.

With regard to the Examiner's comments on claim 3, the Sahami et al. '635 publication does not disclose an indicator comprising a conditional probability of the occurrence of the respective co-occurring term in a document given the appearance in said document of the term in respect of which the probability distribution is calculated, as per claim 3. In contrast, the Sahami et al. '635 publication's use of conditional probability is of "mutual information" (see paragraph [0066]), where "mutual information" is defined as the influence between pairs of attributes assigned to a data set (see paragraphs [0061] to [0064]). As indicated previously, the "attributes" of the data set taught by the Sahami et al. '635 publication are not related in any way to the terms (e.g. words) in a document, and the Sahami et al. '635 publication's "mutual information" is not related in any way to the occurrence of terms in a document. Hence, the features of claim 3 are not disclosed by the Sahami et al. '635 publication.

With regard to the Examiner's comments on claim 4, the Tukey et al. '422 patent teaches the normalization of word overlap counts and document vectors (column 2, lines 28-47). In contrast, claim 4 recites the normalization of "indicators," which, as defined in claim 1 and discussed above, comprise probability distributions relating to the occurrence of terms in documents. Hence, the Tukey et al. '422 patent does not disclose the features of claim 4.

With regard to the Examiner's comments on claims 5 and 6, in the Sahami et al. '635 publication, the phrase "a subset of the set of data" means set of data records that are members of same cluster. This is not the same as the subsets of terms (e.g. words) recited in claim 5. Also, the Sahami et al. '635 publication does not disclose that the subsets are assigned depending on the frequency of occurrence of the term, as required by claim 5. With regard to the Tukey et al. '422 patent's disclosure concerning selecting cluster attractors, as explained above, the Sahami et al. '635 publication does not use cluster attractors and so it would be impossible for one of

ordinary skill in the art to combine the Tukey et al. '422 patent's teachings in this regard with the Sahami et al. '635 publication. Therefore, neither the Sahami et al. '635 publication nor the Tukey et al. '422 patent, either individually or combined, disclose or suggest the features of claims 5 and 6.

With regard to the Examiner's comments on claim 7 and 8, these claims recite entropy thresholds. The Sahami et al. '635 publication does not disclose entropy thresholds – the thresholds disclosed in Table 1 of the Sahami et al. '635 publication are thresholds on "influence score" only. Moreover, as described in relation to claim 1, the Sahami et al. '635 publication does not make any disclosure concerning disclose the entropy of probability distributions relating to the occurrence of terms in documents, which is also included in claims 7 and 8. With regard to the Tukey et al. '422 patent's disclosure concerning cluster attractors, as explained above, the Sahami et al. '635 publication does not use cluster attractors and so it would be impossible for a skilled person to combine the Tukey et al. '422 patent's teachings in this regard with the Sahami et al. '635 publication. Therefore, the Sahami et al. '635 publication and the Tukey et al. '422 patent, either individually or combined, do not disclose or suggest the features of claims 7 and 8.

With regard to the Examiner's comments on claim 9, in the Sahami et al. '635 publication the "frequency information" is computed for a set of attributes of a data set (see claims 1 and 5 of the Sahami et al. '635 publication). The Sahami et al. '635 publication does not disclose frequency ranges, or associating frequency ranges with subsets. In the Tukey et al. '422 patent (column 13, lines 8-16) the phrase "disjoint" relates to a special type of clustering ("hard" clustering) when any data record belongs to one and only one cluster. The Tukey et al. '422 patent does not disclose subsets that are disjoint. Therefore, the Sahami et al. '635 publication and the Tukey et al. '422 patent, either individually or combined, do not disclose or suggest the features of claim 9.

With regard to the Examiner's comments on claim 10, the Sahami et al. '635 publication does not disclose successive frequency ranges being equal to a constant multiplied by the size of the preceding frequency range in order of increasing frequency. Instead, the Sahami et al. '635 publication just discloses frequency vectors (paragraph [0076]). The vectors of increasing length

mentioned in paragraph [0009] of the Sahami et al. '635 publication are part of a discussion of the prior art, which is not used by the Sahami et al. '635 publication and has no relation to the Sahami et al. '635 publication's frequency vectors. Hence, the Sahami et al. '635 publication does not disclose the feature of claim 10.

Similar comments apply to the Examiner's comments on claims 11 and 12, as were made in relation to claims 7 and 8.

Amended claim 15 relates to a clustering method using the method of claim 1 and so similar comments apply as were made in relation to claim 1.

With regard to the Examiner's comments on claim 16 and 17, as indicated above, the Sahami et al. '635 publication does not operate on the terms (words) of a document. Instead, its analysis is performed on "attributes" of the data. Hence, the Sahami et al. '635 publication does not disclose the calculation of probability distributions of the occurrence of terms of each document. The Tukey et al. '422 patent does not disclose the calculation of probability distributions of the occurrence of terms of each document or the comparison of these against probability distributions selected as cluster attractors.

It is at least for these reasons that the cited references fails.

Therefore, it is submitted that independent claim 1 and all the claims depending therefrom are unobvious over the cited prior art of record, whether taken alone or in any combination.

Similarly, it is submitted that independent claims 14 and 18, of similar scope as independent claim 1, are unobvious over the cited prior art of record, whether taken alone or in any combination.

It is therefore respectively submitted that the rejections under 35 U.S.C. 103(a) should be withdrawn.

CONCLUSION

In light of the foregoing, Applicants submit that the application is in condition for allowance. If the Examiner believes the application is not in condition for allowance, Applicants respectfully request that the Examiner call the undersigned.

Respectfully submitted,

THE NATH LAW GROUP

November 17, 2008

THE NATH LAW GROUP
112 South West Street
Alexandria, VA 22314-2891
Tel: 703-548-6284
Fax: 703-683-8396



Jerald L. Meyer
Registration No. 41,194
Derek Richmond
Registration No. 45,771
Jiaxiao Zhang
Registration No. 63,235
Customer No. 20529